HAL open science

New Parameters for Measuring Exposure to Diffusion in Dynamic Networks

Jérémie Dumas, Christophe Crespelle, Eric Fleury

To cite this version:

Jérémie Dumas, Christophe Crespelle, Eric Fleury. New Parameters for Measuring Exposure to Diffusion in Dynamic Networks. LAWDN - Latin-American Workshop on Dynamic Networks, INTECIN - Facultad de Ingeniería (U.B.A.) - I.T.B.A., Nov 2010, Buenos Aires, Argentina. 4 p. inria-00531748

HAL Id: inria-00531748
https://inria.hal.science/inria-00531748

Submitted on 4 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

New Parameters for Measuring Exposure to Diffusion in Dynamic Networks

Jérémie Dumas
ENS de Lyon
Email: jeremie.dumas@ens-lyon.org

Christophe Crespelle
Université Claude Bernard Lyon 1, DNET/INRIA
LIP UMR CNRS 5668, ENS de Lyon
Université de Lyon
Email: christophe.crespelle@inria.fr

Eric Fleury
ENS de Lyon, DNET/INRIA
LIP UMR CNRS 5668
Université de Lyon
Email: eric.fleury@inria.fr

***Abstract*—This paper deals with diffusion processes in dynamic networks. We propose new notions for measuring the exposure of a node to diffusion processes taking place over the network, that is the likelihood of the node to be affected by such processes. We introduce different exposure scores, based either on contacts or on flows in the dynamic network. We study the distribution of these values in the example case of a real-world dynamic network and investigate possible correlations between them. We also discuss computational complexity issues.**

I. Introduction

Diffusion processes are studied in many different contexts: diseases in human groups, information in societies, messages in communication networks. In all these contexts, these processes have a major importance and this is the reason why they have been extensively studied. Usually, these diffusion processes take place over a *network* defined by a set of pairwise *links* between the entities (called *nodes*) of the group where the diffusion takes place. That is, one node u can transmit a certain property $\mathcal{P}$, which we study diffusion of, to another node v only if these two nodes are linked together in the network. In epidemiology, for example [1] [2], these links correspond to contacts between individuals, while in the case of a message to be diffused over a communication network such as the internet, the links are the wires between two computers.

In many cases (e.g. when nodes are mobile), these contacts change along time, giving rise to what is called a *dynamic network*. There is no doubt that the way the links evolve in the network along time has a strong influence on the ability of property $\mathcal{P}$ to diffuse. Moreover, every node of the network has a particular behaviour that may change its ability to be reached by the diffusion of property $\mathcal{P}$. Our purpose is precisely to give quantitative notions able to evaluate the likelihood of a node u to be affected by a diffusion originating from a given subset of nodes in the network, which we call the *exposure* of u.

Our contribution. This paper is a prospective paper. Our main contribution is to define several notions to estimate the exposure of nodes to diffusion processes in the network. We give three different definitions, two of them based on contacts and one based on dynamic flow. We compute these parameters on a publicly available dataset [3] consisting of contacts between participants to the Infocom 2005 conference. We study the distribution of the exposure scores we define within this dynamic network and investigate correlations between them.

General definitions. Throughout the paper, a *dynamic network* will be represented by a sequence of undirected graphs $(G_t = (V, E_t))_{1 \leq t \leq p}$, where the set V of vertices is fixed and the set E_t of edges changes along time. Graphs G_t are called *snapshots* of the dynamic network. We denote by $N_t(u)$ the set of neighbours of $u \in V$ in the graph G_t. And for a subset $S \subseteq V$ of vertices , $N_t(S) = \bigcup_{u \in S} N_t(u)$. The generic property which we study the diffusion of in the dynamic network will be denoted by $\mathcal{P}$.

II. Contact-based exposure

In this section we propose to evaluate exposure of a node based on the contacts between this node and other nodes of the network. More precisely, we aim at estimating the exposure of node u during a time period of length δ ending at time t with respect to a diffusion emanating from a subset S of nodes of the network. We give two different definitions of exposure depending on the contacts taken into account.

A. Direct contacts

We start with the most direct definition of exposure one could think of : it consists in counting contacts between node u and nodes of S occurring within the considered time period.

Definition 1: Let $(G_t = (V, E_t))_{1 \leq t \leq p}$ be a dynamic network and $\delta \in [\![1, p]\!]$. For any $t \geq \delta$, we define the *direct contact exposure* of node $u \in V$ at time t with respect to subset $S \subseteq V$ of nodes over a δ time period, denoted $E_{S,\delta}(u, t)$, as the number of contacts between u and nodes of $S \subseteq V$ within the interval $[\![t - \delta + 1, t]\!]$. Formally,

$$E_{S,\delta}(u, t) = \sum_{i=t-\delta+1}^{t} |S \cap N_i(u)|$$

B. Extended contacts

The main drawback of the above notion is that it does not take into account the possibility of indirect diffusion. Indeed, the property $\mathcal{P}$ whose diffusion is considered can be transferred from nodes of S to vertex u by the mean of an intermediate node $v \notin S$ who itself obtained the property $\mathcal{P}$

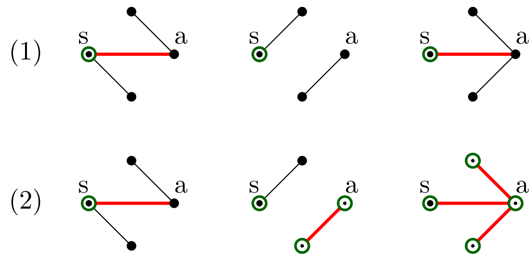

Fig. 1. Example of a dynamic network. (1) direct contacts. (2) extended contacts : the maximal diffusion is represented by circles. For both kinds of contacts, links contributing to the exposure of a appear in bold.

from nodes of S. Thus, all contacts of node u with nodes v that have already been in contact with some node of S may contribute to diffuse property $\mathcal{P}$ to node u. Here we aim at taking this contribution into account in the exposure of node u.

To that purpose, we rely on the notion of *maximal diffusion* which consists in considering at each step the subset $\overline{S}$ of vertices that could possibly have obtained property $\mathcal{P}$ by direct or indirect transmissions (see Figure 1 (2)). In other words, the maximal diffusion is a virtual diffusion in which all contacts involving a node v having property $\mathcal{P}$ result in a transmission. We adopt the following notation.

Definition 2: Let $(G_t = (V, E_t))_{1\leq t\leq p}$ be a dynamic network and $a \in [\![1, p]\!]$. We denote by $\overline{S}_{a,t}$ the set of vertices obtained at time t by maximal diffusion starting from S at time a. Formally,

$$\overline{S}_{a,t} = \begin{cases} \emptyset & \text{if } t < a \\ S & \text{if } t = a \\ \overline{S}_{a,t-1} \cup N_{t-1}(\overline{S}_{a,t-1}) & \text{otherwise} \end{cases}$$

Thanks to the notion of maximal diffusion, we define the *extended contact exposure* of node v at time t, over a δ time period, as the number of contacts, during this time period, between v and vertices of $\overline{S}$ resulting from the maximal diffusion staring from S at time $t - \delta + 1$.

Definition 3: Let $(G_t = (V, E_t))_{1\leq t\leq p}$ be a dynamic network and $\delta \in [\![1, p]\!]$. For any $t \geq \delta$, we define the *extended contact exposure* of node $u \in V$ at time t with respect to subset $S \subseteq V$ of nodes over a δ time period, denoted $E^*_{S,\delta}(u, t)$, as the number of contacts between u and nodes of $\overline{S}_{t-\delta+1,i}$ for $i \in [\![t - \delta + 1, t]\!]$. Formally,

$$E^*_{S,\delta}(u, t) = \sum_{i=t-\delta+1}^{t} \left|\overline{S}_{t-\delta+1,i} \cap N_i(u)\right|$$

C. Time complexity

Let $n = |V|$ be the number of vertices in the network and $m_{max} = \max\{|E_t|, t\}$ the maximum number of edges in a snapshot. The time complexity of computing $E^*_{S,\delta}(u, t)$ is $\mathcal{O}((n+m_{max})\delta)$. Indeed, one can store the set $\overline{S}_{t-\delta+1,i}$ using an array of size n, which can be updated in time $\mathcal{O}(n+m_{max})$ at each time step with a simple graph traversal of G_i.

III. Flow-based exposure

In this section, we give another possible definition of exposure based on flows in a dynamic network. The previous definitions aimed at measuring the exposure of node u by the number of contacts likely to diffuse property $\mathcal{P}$ to u. Here, we evaluate the exposure of node u by assigning an amount of property $\mathcal{P}$ to each node having this property and by considering the maximal amount of property $\mathcal{P}$ that can be transferred to u within a given time period. That is, the exposure is defined as the maximal flow that can transit to u in the dynamic network.

A. Definitions

Before defining *flow exposure* we have to clarify what is a flow in a dynamic network. In the definition we adopt, at each time step: i) the outgoing flow from each vertex v is allowed to traverse only edges incident to v; ii) the flow is authorised to wait on a node without flowing to its neighbours. In other words, paths carrying the flow are composed by a succession of "jump" and "wait".

More formally, let $G_t = (V, E_t)$ be a dynamic network, with c_t a sequence of capacity functions such that $\forall x, y, t, xy \notin E_t$ and $x \neq y \Rightarrow c_t(x, y) = 0$. Let $s, a \in V$ and $t_0, t_1 \in [\![1, p]\!]$, we define the dynamic flow between (s, t_0) and (a, t_1) in G_t as follows.

Definition 4 (Dynamic flow): We say that $(\varphi_t)_{t_0\leq t\leq t_1}$ is a *dynamic flow* of source (s, t_0) and sink (a, t_1) in the network $(G_t)_t$ if and only if it satisfies all the following properties:

1) **Positivity**: $\forall x, y, t, \varphi_t(x, y) \geq 0$
2) **Capacity constraints**: $\forall x, y, t, \varphi_t(x, y) \leq c_t(x, y)$
3) **Conservation**: $\forall x, t, \sum_y \varphi_t(y, x) = \sum_y \varphi_{t+1}(x, y)$
4) **Limit condition**:
   a) $\forall x, y \in V, x \neq s \Rightarrow \varphi_{t_0}(x, y) = 0$
   b) $\forall x, y \in V, y \neq a \Rightarrow \varphi_{t_1}(x, y) = 0$

The *value of the flow*, denoted $|\varphi|$, is defined as the total amount of flow travelling from (s, t_0) to (a, t_1), namely $|\varphi| = \sum_y \varphi_{t_0}(s, y)$.

In the following, we always consider the same sequence of capacity functions c_t, which we call the *canonical capacity sequence*, and which is defined by $c_t(x, y) = 1$ for all $xy \in E_t$ and $c_t(x, x) = +\infty$ for all x, t. This means that each edge of the dynamic network can carry at most one unit of flow, while an arbitrary amount of flow can wait on any node of the network at each time step. Using this capacity functions, we can define the *flow exposure* of a node in the network.

Definition 5 (flow exposure): Let $(G_t = (V, E_t))_{1\leq t\leq p}$ be a dynamic network and $\delta \in [\![1, p]\!]$. For any $t \geq \delta$, we define the *flow exposure* of node $u \in V$ at time t with respect to node $s \in V$ over a δ time period, denoted $F_{s,\delta}(u, t)$, as the the maximum value of a flow from $(s, t - \delta + 1)$ to (u, t).

Remark 1: Though the above definition only stands for flow originating from a single vertex, it can be easily extended to a subset $S \subseteq V$ of vertices. Indeed, one can simulate a flow originating from S by adding a dummy vertex in the network at time $t - \delta$, connected to the vertices of S.

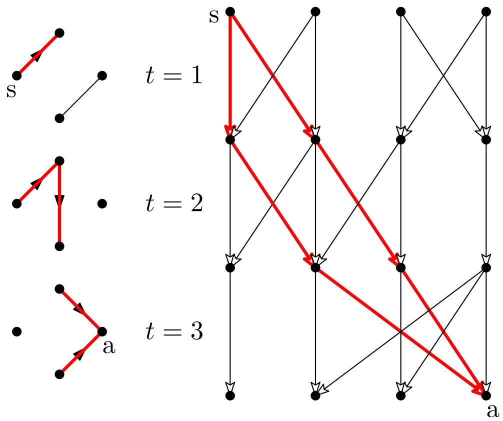

Fig. 2. Example of a dynamic network (left) with its transition graph (right). In bold, the dynamic flow from $(s, 1)$ to $(a, 3)$.

Note that the value of the flow exposure is always between the direct contact exposure and the extended contact exposure:

$$\forall S \subseteq V, \forall u \in V, \forall t, \forall \delta, E_{S,\delta}(u,t) \leq F_{S,\delta}(u,t) \leq E^*_{S,\delta}(u,t)$$

Indeed, a unit of flow can be transmitted to u at each contact with a node of S, which gives the lower bound, and each amount of flow arriving to u must come from a node reached by the maximal diffusion, which gives the upper bound.

B. Transition graph

The dynamic flow defined above can be transformed into a classical flow on a static graph which we call the *transition graph* and which was first introduced in [4].

Definition 6 (Transition graph): Let $(G_t)_t$ be a dynamic network. We denote by $\mathcal{G} = (V^{\mathrm{T}}, E^{\mathrm{T}})$ its transition graph, which is the directed graph defined as follows:

- $V^{\mathrm{T}} = \{(x,t), x \in V, 0 \leq t \leq p\}$
- $E^{\mathrm{T}} = \{((x,t-1),(y,t)), x = y$ or $(x,y) \in E_t, 1 \leq t \leq p\}$

An example of a dynamic network and its transition graph is given in Figure 2. It is not difficult to see that copying the capacities of the dynamic network to its transition graph, each dynamic flow correspond to a classical flow of same value on the transition graph, and vice-versa. We use this feature in order to compute the value of a maximal dynamic flow using classical max-flow algorithms on the corresponding transition graph.

C. Time complexity

As we said previously, the dynamic flow can be expressed as a flow in $\mathcal{G}$. Thus, in order to compute the maximum value of a dynamic flow in $(G_t)_t$, we build the corresponding subgraph of $\mathcal{G}$, to which we apply the Edmonds-Karp algorithm [5] in order to find a maximum flow. Unfortunately, this method is slow when the subgraph of $\mathcal{G}$ is very large.

Indeed, if $(G_t)_t$ is a dynamic network on n vertices having at most m_{max} edges per snapshot, then, for a given set of parameters s, a, t and δ, the subgraph $H \subseteq \mathcal{G}$ corresponding to the sequence $(G_i)_{t-\delta+1 \leq i \leq t}$ has exactly $n \times (\delta+1)$ vertices, and at most $(n + 2m_{max}) \times \delta$ edges (n for vertical edges, $2m_{max}$ for adjacent vertices).

Eventually, the time complexity of the whole algorithm is $\mathcal{O}(NM^2)$, where $N = n\delta$ and $M = (n + 2m_{max})\delta$, which leads to an $\mathcal{O}((n^3 + nm^2_{max})\delta^3)$ bound.

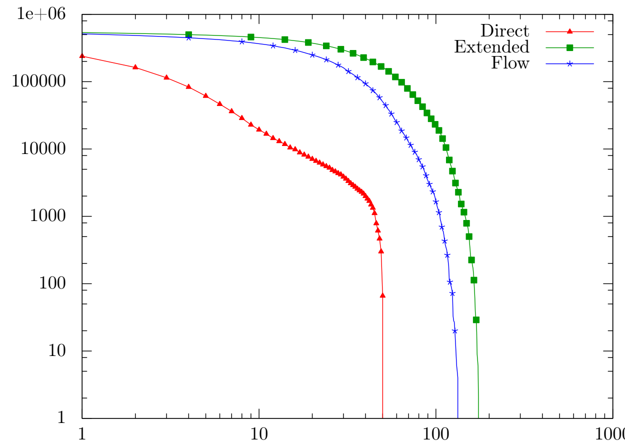

Fig. 3. Inverse cumulative PDF of the three exposure scores plotted in log-log scale. Red triangles for direct contact exposure, green squares for extended contact exposure and blue stars for flow exposure.

IV. Experimentations

We computed the scores of exposure corresponding to the three above-mentioned definitions of exposure on a real-world network using contact data collected during the Infocom 2005 conference [3]. A more thorough study of the classical properties of this network can be found in [6]. This network involves 41 nodes and, after aggregation of the contact data every 300s, the series of snapshots G_t describing its dynamics is composed of 848 graphs. We fixed the time period δ on which we evaluate exposure to $\delta = 50$ graphs (i.e. 15 000s). The question of fixing accurately this value is an important issue in itself but is far beyond the scope of this paper. One should keep in mind that results presented here holds for this particular choice but may not for others. Since there is no diffusion in our data set, there is no clear choice of set S which we should compute exposure with respect to. We chose to compute exposure with respect to every singleton $S = \{v\}$ where $v \in V$. That is, we obtain a score for any triplet (u, v, t) which is the exposure of u with respect to v at time t.

Figure 3 shows the inverse cumulative Probability Distribution Functions (PDF) of the three different scores of exposure for all triplets (u, v, t). The curve of the direct contact exposure seems to be close to a straight line in its left part, revealing a possible power-law nature. The curve also presents a clear cut-off at 50 which is the value of δ. Clearly, testing different values of δ would be highly desirable in this case.

For the notion of exposure based on extended contacts as well as flow, the situation is very different: the two distributions appear to be more homogeneous. This fact is interesting as it suggests that the direct contacts and extended contacts are

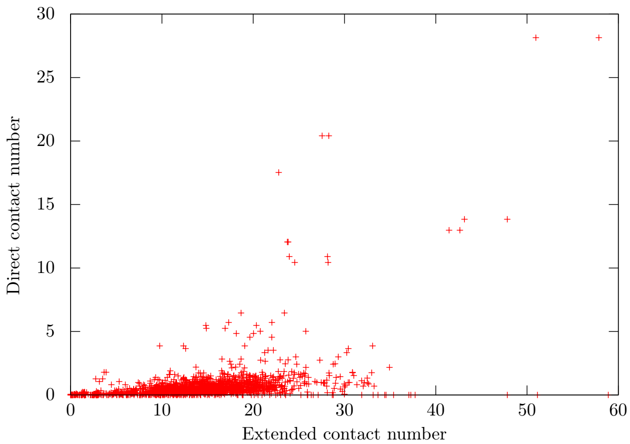

Fig. 4. Correlation between direct (ordinate) and extended (abscissa) contact exposure for all couples (u, v) of vertices in the network (averaged over time). Correlation coefficient $r = 0.407$.

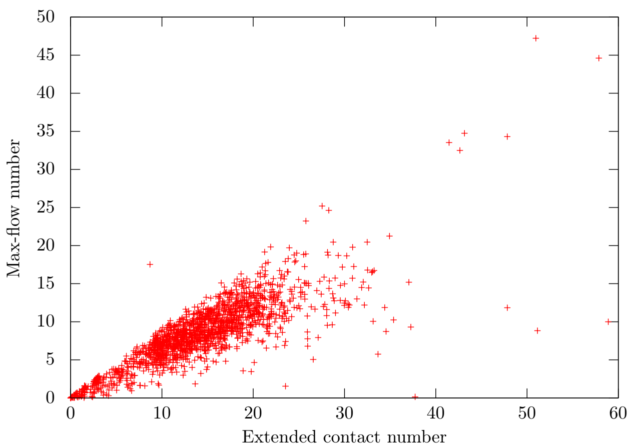

Fig. 5. Correlation between max-flow exposure (ordinate) and extended contact exposure (abscissa) for all couples (u, v) of vertices in the network (averaged over time). Correlation coefficient $r = 0.823$.

quite different notions. On the other hand, directed contacts and flow seems to give rise to comparable scores of exposure.

In order to investigate further these questions, we present on Figure 4 a scatter plot where each point (x, y) correspond to a couple (u, v) of vertices in the network: x is the extended contact exposure of u with respect to v averaged over time and y is the average direct contact exposure. One can see that extended contact exposure is usually much greater than direct contact exposure, and it makes uneasy to appreciate whether there is a correlation between the two values. To that purpose, we also computed their linear correlation coefficient, which is 0.407. This rather low value confirms that the two types of contacts, direct and extended, are pretty different.

Similarly, Figure 5 shows a scatter plot where x is the average extended contact exposure and y is the average flow exposure, for all couples (u, v). It appears clearly that most points are closed to a straight line. This is confirmed by the linear correlation coefficient which is 0.823, indicating that these two parameters are fairly linked, in this case. This would be very interesting in practice since extended contact might then be used to approximate flow, which is much more difficult to compute (see the previous complexity sections).

These observations have to be tempered with the fact that they may be sensitive to the value of parameter δ. In particular, a too large value of γ may result in a maximal diffusion that reaches all nodes of the network, thereby reducing the notion of extended contacts to the number of contacts of the node in the network. Considering different values of δ is crucial and constitutes one of the main perspectives of our work.

V. CONCLUSION

We introduced three different parameters, called exposure, for estimating the likelihood of a node u to be affected by a diffusion process in a dynamic network. Two of them are based on the contacts of node u while the third one lean on the notion of maximal flow in a dynamic network.

We computed the values of these parameters on a real-world network, the results ask for several questions. Is there a correlation between extended contact exposure and flow exposure? Are direct contacts and extended contacts always poorly correlated? Another question of particular interest is the choice of the length δ of the time period on which we evaluate exposure. This parameter could have significant impact on the results and choosing it appropriately is extremely important in practice.

Finally, since the goal of exposure is to evaluate the likelihood of a node to be influenced by diffusion processes taking place over the network, the most promising perspective of our work is to use exposure to analyse real-world diffusion data in dynamic networks. Then, we would be able to determine whether these three notions are able to explain the diffusion dynamics and to predict which nodes will be affected in the future. Moreover, this would allow to compare these three notions between them and to find out which are more accurate.

REFERENCES

[1] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley, 2000.

[2] R. Pastor-Satorras and A. Vespignani, “Epidemic dynamics and endemic states in complex networks,” *Phys. Rev. E*, vol. 63, no. 6, p. 066117, May 2001.

[3] D. Kotz and T. Henderson, “CRAWDAD: A Community Resource for Archiving Wireless Data at Dartmouth,” *IEEE Pervasive Computing*, vol. 4, no. 4, pp. 12–14, 2005.

[4] A. Ferreira and L. Viennot, “A note on models, algorithms, and data structures for dynamic communication networks,” INRIA, Technical Report RR-4403, 2002.

[5] J. Edmonds and R. M. Karp, “Theoretical improvements in algorithmic efficiency for network flow problems,” *J. ACM*, vol. 19, no. 2, pp. 248–264, 1972.

[6] A. Scherrer, P. Borgnat, E. Fleury, J.-L. Guillaume, and C. Robardet, “Description and simulation of dynamic mobility networks,” *Computer Networks*, vol. 52, no. 15, pp. 2842–2858, 2008.